A MODEL FOR BELIEF REVISION IN A MULTI-AGENT ENVIRONMENT

Aldo Franco Dragoni

University of Ancona, Istituto di Informatica
via Brecce Bianche 60131 Ancona
ITALY
tel. +39-71-2204832; fax +39-71-2204474
dragon@anvax2.cineca.it

In modeling the knowledge processing structure of an Agent in a Multi-Agent world it becomes necessary to enlarge the traditional concept of Belief Revision. For detecting contradictions and identifying their sources it is sufficient to maintain informations about *what* has been told; but to "solve" a contradiction it is necessary to keep informations about *who* said it or, in general, about the source where that knowledge came from. We can take as certain the fact that an agent gave an information, but we can take the given information only as a revisable assumption. The Belief Revision system can't leave the sources of the informations out of consideration because of their relevance in giving the additional notion of "strength of belief" [Galliers 89]. In fact, the reliability of the source affects the credibility of the information and vice-versa. It is necessary to develop systems that deal with couples <assumption, source of the assumption>. In [Dragoni 91] we've proposed a system that moves in this direction. Here we give a short description of that system. In the first two parts we describe the agent's knowledge processing structure using a particular characterization of the "Assumption Based Belief Revision" concept; in part three we outline the project of an embedded device that enables the overall system to deal with couples <assumption, source of the assumption> in an rather anthropomorphic manner.

1 Introduction

By "Belief Revision" we mean the process of detecting contradictions, identifying the assumptions from which they came out and *readjusting the* *knowledge base to remove the contradictions*. Beliefs are assumed to be expressed as sentences of first order logic stored in the agent's memory. Belief sets are not assumed to be closed under logical consequences as in [Gardenfors 90] (i.e. if K is a belief set then, generally, K≠Th(K), where Th(K) denotes all the sentences derivable by a complete first-order theorem prover from K). There are two kinds of sentences: those actually introduced as assumptions and those actually deductively derived as logical consequences of the assumptions (respectively *evidences* and *conclusions* in [Post 90]). We need an *Assumption Based Truth Maintenance System* [De Kleer 86] and we adopt the following modified version of the *Supported Wff* of Martins and Shapiro [Mar-Shap 86] (the rationalities for the multi-agent topic are in part two):

$$SW=<A,OS_O,OS_1,..,OS_n,OS_E,RS>$$

where A is an F.O.L sentence; among the assumptions really used in the derivation of A OS_O contains those whose source is an *observation*, OS_i contains those whose source is a *communication* received from the agent i and OS_E contains those introduced hypothetically by the agent himself; OS = $OS_O \cup OS_1 \cup .. \cup OS_n \cup OS_E$ is the *Origin Set* of A; RS is the *Restriction Set*; it contains all the sets of assumptions that joined with the OS produce a strongly-inconsistent set (see below). Each SW has a distinct identifier.

An *assumption* is an SW whose OS refers (only) to itself. We define *contradiction* a couple of SWs $<A,OS_1,RS_1>$ and $<\neg A,OS_2,RS_2>$. The set $OS_1 \cup OS_2$ from which has been derived the contradiction is defined to be a *strongly-inconsistent* set. A *weakly-consistent* set is a not strongly-inconsistent one. We suppose the presence of a real-time working Reasoner (see below), comprehensive of a first-order resolution (or natural deduction) subsystem. The Reasoner clocks the (simulated) time by means of two distinct class of events: the assumption of a new SW and the deduction of a new SW. No more than one assumption can be introduced at a time, no more than one SW can be derived at a time, and no sentence can ever be erased from the memory (as the reasoning goes on, an SW can be changed into another one with different OS and RS but with the same sentence).

The central idea of an Assumption Based Belief Revision System is that of *preventing the system from reasoning upon a set of assumptions already marked as a strongly-inconsistent one*. The main problem is that of defining criteria *to select the best set of assumption to reason with* among the many possible outcomes of the ATMS.

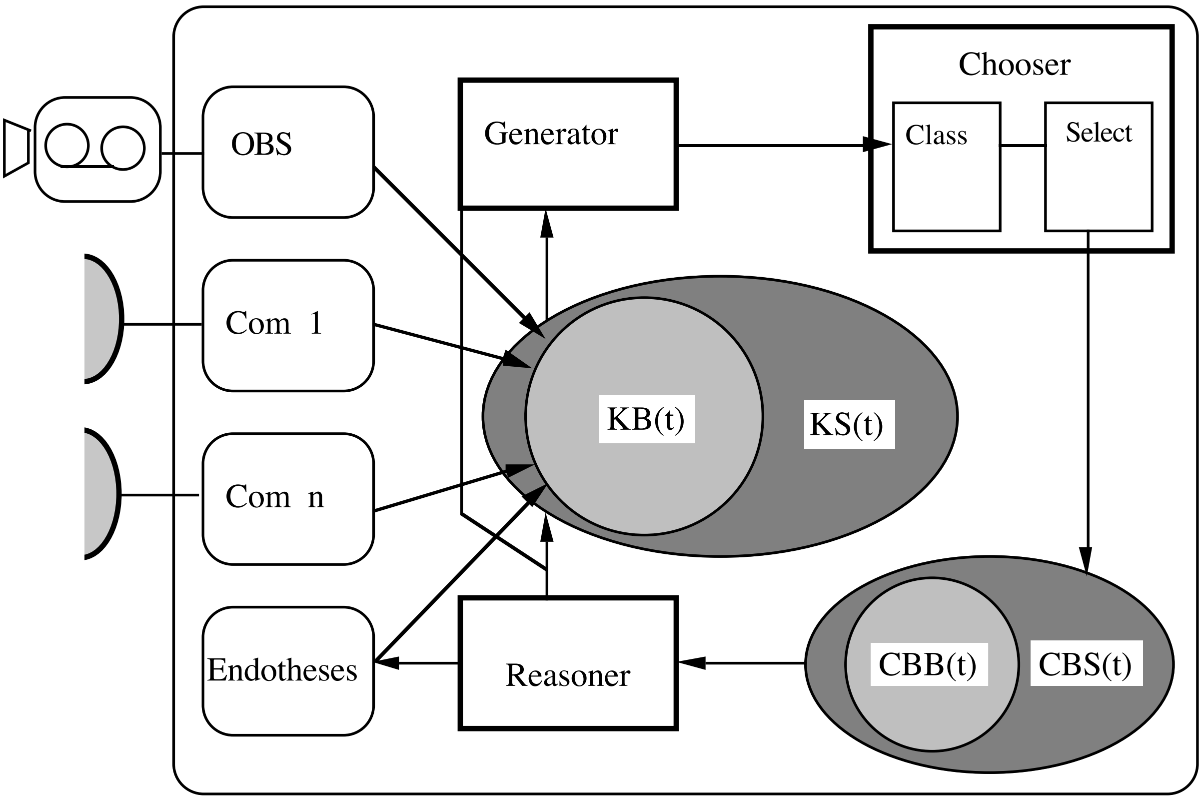
The *Knowledge Base KB(t)* is a mapping from time to sets of sentences; these sentences are to be taken as the assumptions introduced by the Reasoner up to the time t. The *Knowledge Space KS(t)* is a mapping from time to sets of sentences too, but these sentences are to be taken as those deductively derived from KB(t) by the Reasoner up to the time t. For each t∈T, a *Belief Base $BB_i(t)$* is a subset of KB(t) such that:

1. it is weakly-consistent,

2. it is *maximal* with respect to KB(t) (if augmented with whatever else assumption of KB(t) it becomes a strongly-inconsistent set).
The *Belief Space* $BS_{BBi(t)}(t)$ is the set of all the sentences derived from BB(t) up to t. It follows from the definitions that, for every t∈T:
1. KB(t) ⊆ KS(t), (= if the reasoner has not yet derived any sentence),
2. KB(t) ⊆ KB(t+1) (= when time is clocked by a deduction),
3. KS(t) ⊆ KS(t+1) (= when time is clocked by an assumption),
4. KS(t) ⊆ Th(KB(t)) (because of the monotonicity of Th).

2 The Belief Revision System

We use an ATMS to model the specific Belief Revision process of a *single* Agent living in a Multi-Agent world; this is not a Distributed Assumption Truth Maintenance System as in [Mason 89]. With reference to the picture we sketch here the entire system.



2.1 The REASONER

Its essential task is to clock simulated time providing the assumption of a new SW in KB(t) (these assumptions will be named "Endotheses" below), or the deduction of a new SW in KS(t). The first activity is intended to model forms of plausible reasoning [Davis 90] (abduction, induction, nonmonotonic reasoning [Gen-Nils 87] etc.); the second activity is intended to model the limited deductive ability of a real reasoning agent. KB(t)⊆KS(t) because assumptions are logical consequences of themselves. No sentences will ever be removed from KS(t). We call *Current Belief Base* CBB(t) the particular Belief Base chosen by the Chooser (defined below) as the preferred one. We call *Current Belief Space* the set $CBS(t)=BS_{CBB(t)}(t)$. The intended meaning for CBS(t) is to be the most believable and maximal piece of knowledge actually available for the reasoning agent. Probably, for best results, it would be preferable to limit at CBS(t) (instead of at the entire set of Belief Bases in KS(t)) the input of the Reasoner (as depicted in the picture) but we see no serious advantages to be so drastic; simply it seems to be curious to elect a Belief Base as the preferable one and then to reason upon another one. In our implementation [Pioli 91] the reasoning task is carried out by the user on a recent version of FOL by Weyhrauch [Giunchiglia 91].

2.2 The Belief Bases' GENERATOR

We belong to the group of those that think Consistency valuable for Mathematical Logic but destructive in modelling CommonSense Reasoners [Minsky 81]. Nevertheless we think Consistency cannot be disregarded in a lot of scientific or technical areas (medicine, engineering, law and so on), hence, for our purpose, it would be very desirable for the set of agent's belief base CBB(t) to be consistent. Unfortunately, practical FOL-based systems have to restrict themselves to consider only limited forms of Consistency because of the indecidibility of the validity problem. Previously defined Weak-Consistency is our limited form of Consistency. Note that while Consistency is a property of sets of sentences, Weak-Consistency is a property of sets of SWs. It is possible to have two sets of SWs with the same sentences where the first is weakly-consistent while the second is not. It seems also desirable for an agent to use as more informations as possible in its reasonings; we think that a theory of the AIDS with the statement "speed(light,300.000)" is preferable than the same theory without that statement because we are not absolutely sure that light will never have something to do with AIDS. Hence our choice to impose maximality for CBB(t). Notice that this maximality is intended with respect to all the assumptions in KB(t) and not, as usually, with respect of all the sentences of the Language; this is because we give no importance to the sentences not already introduced in the memory. Each event is a clock pulse for the Belief Base Generator. It searches all over KS(t) for a contradiction. If it succeeds it records the discovery of the strongly-inconsistent set and redefines the Situation S(t) of all the Belief Bases in KB(t) (see the *Updating Restriction Set* rule in [Mar-Shap 87] for details).

2.3 The CHOOSER

After the discovery of a new contradiction the agent is in a position to revision its beliefs. It is not the case to select which belief is to be thrown away to remove the contradiction, but, quite more generally, to choose which is the *new preferred Belief Base* among them in S(t). This is what we mean by "Belief Revision" and this is the task of the Chooser; it is an appropriate machine that takes S(t) as input and gives the new preferred Belief Base CBB(t) as output. We've found useful to think the Chooser as the cascade of two components: the Classifier and the Selector. The Classifier takes KB(t) as input an gives as output the list of all the assumptions in KB(t) ordered according to some specific criteria. The Selector takes as input the list passed from the Classifier and the situation S(t) passed by the Generator and gives as output CBB(t). In order to produce the list of the assumptions, the Classifier needs not only some credibility-importance criteria to judge them but also a strategy to manage those criteria. These could in fact be used at least in two different ways: they could be sorted according to their importance (the importance of the criteria themselves) and used in cascade as selective filters on the assumptions, or it could be assigned a weight to each one of them according to their importance and then they could be used as tests score on the assumptions, reporting the degree with which they are satisfied. The first kind of strategy is a particular case of the second one; it seems preferable because of its simplicity and its apparent closeness to the human usage; the followings could be its steps:

```
begin
sort the n criteria in criteria[i] by their importance;
KB:=KB(t); stack:=empty;
while not_empty KB do
    begin
    i:=1;
    repeat
        if there are assumptions in KB verifying criteria[i]
            then select only these
            else select all the assumptions in KB
        i:=i+1
    until i=n;
    pop an assumption from KB;
    push that assumption in stack
    end
end
```

At the halt of this algorithm, `stack` contains the ordered list of assumptions which is the output of the Classifier. The Selector takes the stack and S(t) as input and gives one of the Belief Bases in S(t) as output. Many possible strategies are based on the following procedure:

```
begin
Select all the Belief Bases in S(t);
repeat
Pop an assumption from stack;
if there exist some Belief Bases containing that assumption
    then Select only these
    else Select all the Belief Bases
until there is only one Belief Bases
end.
```

The last Belief Bases is the preferred one.

This system shows both *foundational* and *coherence* nature. From [Galliers 89]: "Foundation theory considers new beliefs are only to be added on the basis of other justified beliefs, and beliefs no longer justified are abandoned", this is the case of beliefs corresponding to the deductively derived sentences which are in CBS(t) and are added only on the basis of the set of assumptions CBB(t); "Coherence theory represents a conservatism whereby justification is only a requisite condition of believing if there is a special reason to doubt a belief", this is the case of beliefs corresponding to the assumptions whose permanence in CBB(t) is only due to their being not strongly-inconsistent with the others in the same Belief Base; the assumptions in a Belief Base are there because there isn't a valid reason for their not being there.

3 The Multi-Agent setting

We distinguish three kinds of sources:

- *perception* (typically vision) gives a first direct information about the state of the world (objects' and agents' spatial positions etc.) and about spatial events or occurring actions;
- *communication*: each agent is able to exchange informations employing a certain physical channel and appropriate communication protocols; agents are not necessarily sincere nor competent ones;
- *reflection*: for the sake of realism we admit the presence of assumptions engendered by some forms of hypothetical reasoning internal to the agent; we call them *Endothesis* and we discuss below a problem with them.

In addition we see the presence of "a priori" assumptions that, for our purpose, could be though as innate to the agent; they represent, typically, the rules (causal or not) of the knowledge domain under consideration but we think them as not removable and, therefore, not assumptions at all.

It is also fundamental to maintain distinction among communications received from different agents. This leads to the previously introduced version of the Supported Wff.

3.1 Some Criteria for judging assumptions in a Multi-Agent setting

The following is a proposal list.

1. Assumptions derived from observation are stronger then those derived from communication. Observation is taken as a sort of Super-Agent.

1b. Assumptions derived from communication which are in contrast with sets of assumptions all derived from observation have no strength at all.

2. The sources multiplicity confirms the assumption.

3. The more the conflicts with other assumptions, the weaker the assumption.

4. The OSs of SWs with the same wff confirm each others because of their mutual coherence.

Two criteria modelling psychological attitudes.

5. *Belief Conservativeness*: it is stronger the assumption supporting more SWs.

6. *Goals Conservativeness*: it is stronger the assumption supporting more goals (assuming a planner working on the system).

It is important to consider also the reliability of the agents.

7. The less reliable the agent who made a communication, the weaker the assumption derived from it . We could estimate the agent's reliability by:

- Self-Inconsistency (he made communications mutually inconsistent)
- average of Inconsistency of the assumptions derived from communications received from that agent with respect of all the other assumptions derived from observations in KB(t) (we choose to not consider the conflicts with other assumptions in order not to punish competency).

We emphasize the importance of the 4^{th} criterion in giving a prize to coherence. Let $Th(K_t)$ represent a scientific theory based on a set of assumptions K_t not derived from observation and let K_o be a set of assumptions all derived from observation; if there is a couple of SWs with the same sentence in KS(t), named $\mathbb{S}$ and $\mathbb{T}$, $OS(\mathbb{S})=K_t$ and $OS(\mathbb{T})=K_o$, then K_o could be intended to represent the experimental evidences of the scientific theory so that it is justified its reinforcement over K_t. The 7^{th} criterion could be seen as a preprocessor that gives a weight to each assumption derived from communication. The criteria 1-4 should be able to manage these weights.

3.2 The ENDOTHESES

Realistic situations require much complex treatments. Among other considerations, we endorse the need for assumptions "internal" to the agent; we think them as auxiliary beliefs, functional to the reasoning process which is going on. That's the rationality for the Endotheses. They could be the result of the application of some sort of plausible inference rules, it may be "induction" (roughly from α and β infer $\alpha \rightarrow \beta$) or "abduction" (roughly from β and $\alpha \rightarrow \beta$ infer α) or a nonmonotonic default rule. These Endotheses are treated normally by the Generator; that is, an Endothesis α belongs to every Belief Base not containing a subset that is strong-inconsistent with α. This implies that an Endothesis α hypothetically derived from a set of assumptions $\{\alpha_1,..,\alpha_i\}$ in a Belief Base, can as well belong to other Belief Bases not containing $\{\alpha_1,..,\alpha_i\}$. This could seem strange but it is in accord with the Principle of Positive Undermining [Harman 86]: the lack of justification is not a good reason to remove a belief; we think that this principle is more appropriate for beliefs plausibly derived from a set of assumptions than it is for beliefs which are logical consequences of the set of assumptions. The real problem with the Endotheses is that it isn't clear how they are to be treated by the Chooser; because of the arbitrarity with which they can be introduced we can't fix criteria for the Chooser to process them. The problem is that their introduction is again a casual element in the reasoning story and we are no more favourably disposed towards such casualness. We've been well disposed towards the casualness inherent the story of the introduction of assumptions *from the outside* of the agent (by communication and observation). They are regarded as "interrupts" to be processed, and the change of the deductive theory following the arrive of one of them is justified by the real change of the agent's cognitive state. We've been not so well disposed towards the casualness inherent the story of the derivations of new SWs in KS(t). However, if the Inconsistency of a set is revealed in its being strong-inconsistent it depends on the story of the deductions made by the Reasoner. But Strong-Inconsistency itself does not depend on that story. So, if the property that defines a Belief Base is Weak-Consistency (not Consistency) then we have nothing to worry about the casualness of the deductions.

But now, we are not well disposed towards the casualness inherent the story of the introduction of assumptions *from the inside* of the agent (the Endotheses). It could be objected that these events too could be regarded as interrupts changing the agent's cognitive state and we've just accepted the fact that casual elements internal to the Reasoner activity (deductions) can affect the agent's cognitive state. However, our resolution is to reduce this casualness simply *giving the lowest importance to the Endotheses when subjected to the Chooser's processing*.

4 Conclusions

This paper deals with the concept of Assumption Based Belief Revision in a Multi-Agent environment, that is how to consider also the *sources* of the information (i. e. who gave it) in the general belief revision process. We have briefly presented:

a) a rather innovative general framework for assumption based Belief Revision

b) some abstract criteria to deal with an agent's knowledge base built upon observations, internal hypotheses and several other agent's informative contributions.

The former topic is based on

- choosing a new preferred Belief Base versus removing the beliefs causing the contradiction,
- achieving our defined Weak-Consistency versus achieving Consistency

The latter topic covers:

- the definition of very general criteria to associate each agent advise (or agent himself) with an implicit credibility factor
- the discussion of a criterion to judge the (our defined) Endotheses, that is assumptions derived internally to the agent
- the discussion of strategies that use these criteria to compute the new preferred Belief Base

The overall system exhibits an enviably anthropomorphic behaviour.

4.1 What is missing

This research belongs to a Multi-Agent planning project, but several examples show that the system fits as well in various fields as police investigations or detective stories. The system could also be seen as a module in expert systems (regarding not well established knowledge) built upon multiple experts, eventually contrasting, contributions. The real limitation of this approach is that it doesn't reason about *why* an agent gave an information; we don't take explicitly in count the *intention* [Cohen 90] of the agents or their dependence relations [Castelfranchi 91] as useful elements to judge their utterances. The key element is only consistency with the perception. It could be too little for some purposes.

4.2 Acknowledgements

I want to express my thanks to Mauro Di Manzo for his comments and suggestions on early longer versions of this paper. This work has been supported by the National Resource Council (C.N.R.), "Progetto Finalizzato Robotica", contract n°90.00350.PF67.

References

[Castelfranchi 91]: Cristiano Castelfranchi & Maria Miceli & Amedeo Cesta, Dependence Relations among Autonomuos Agent, in this proceedings.

[Giunchiglia 91]: F. Giunchiglia and P. Traverso, GETFOL User Manual version 1, Technical Report 9107-01, DIST, Univ. of Genoa, via Opera Pia 11A, 16145 Genova, Italy, July 1991.

[Cohen 90]: Philip Cohen & Jerry Morgan & Martha Pollack, *Intentions in Communication*, The MIT Press, Cambridge, Mass., 1990.
[Davis 90]: Ernest Davis, *Representation of Commonsense Knowledge*, Morgan Kaufmann Publisher, san Mateo, Calif, 1990
[De Kleer 86]: Johan de Kleer, An Assumption Based Truth Maintenance System, *Artificial Intelligence* 28, pp. 127-162, 1986.
[Dragoni 91] Aldo Franco Dragoni, A Model for Belief Revision in a Multi-Agent Environment, Tech. Rep., Computer Sciences Inst., University of Ancona (Italy) 1991.
[Galliers 89]: Julia Rose Galliers, Modelling Autonomous Belief Revision in Dialogue, Tech Rep. Cambridge University Comp. Lab., Cambridge (England), 1989.
[Gardenfors 90]: Peter Gardenfors, Belief Revision and Nonmonotonic Logic: Two Sides of the Same Coin?, *Proceedings of the 9th European Confference on Artificial Intelligence*, L. C. Aiello ed., Pitman Publishing, London, 1990.
[Gen-Nils 87]: Michael Genesereth, Nils Nilsson, *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann Publisher, san Mateo, Calif, 1987
[Giunchiglia 91]: F. Giunchiglia and P. Traverso, GETFOL User Manual version 1, Technical Report 9107-01, DIST, Univ. of Genoa, via Opera Pia 11A, 16145 Genova, Italy, July 1991.
[Harmann 86]: G. Harman, Change in View - *Principles in Reasoning*, The MIT Press, Cambridge, Mass., 1986.
[Mar-Shap 86]: Joao P. Martins, Stuart C. Shapiro, Theoretical Foundations of Belief Revision, in: J.Y. Halpern ed., *Theoretical Aspects of Reasoning about Knowledge*, Morgan Kaufmann, Los Altos, CA, 1986.
[Mar-Shap 88]: Joao P. Martins, Stuart C. Shapiro, A Model for Belief Revision, *Artificial Intelligence* 35 (1), pp. 25-79, 1988.
[Mason 89]: Cindy L. Mason and Rowland R. Johnson, DATMS: A Framework for Distributed Assumption Based Reasoning, in L. Gasser and M. N. Huhns eds., Distributed Artificial Intelligence 2 (Pitman/Morgan Kaufmann, London, pp 293-318, 1989.
[Pioli 91]: M. Pioli, Revisione delle Conoscenze in ambiente Multi-Agente: Manuale dell'ATMS, Tech. Rep. Istituto di Informatica, Univ. Ancona, via Brecce Bianche, 60100, Ancona, 1991.
[Post 90]: Stephen Post and Andrew P. Sage, An Overview of Automated Reasoning, *IEEE Transaction of Systems, Man, and Cybernetics,* vol 20 n°1, 1990.